

BENCHMARKS

VOLUME 3 NUMBER 3

NORTH TEXAS STATE UNIVERSITY

MARCH/APRIL 1982

*BENCHMARKS was produced using MUSIC/SCRIPT

TABLE OF CONTENTS

1.	PART III OF USER'S GUIDE AVAILABLE IN THE UNIVERSITY STORE	3
2.	TELL US WHAT YOU WANT	3
3.	DATA ENTRY INTO MUSIC FILES	3
4.	800 BPI TAPES REQUIRE SPECIAL SCHEDULING	4
5.	800 BPI PROCESSING COMING TO AN END	4
6.	SPACE ON ACADEMIC DISK PACKS RUNNING DANGEROUSLY LOW . .	5
7.	USING THE COMPUTER FOR RESEARCH: PART II	7
8.	SPSS	15
	Release 9.1 Installed	15
	Known Errors In SPSS Release 9.1	16
9.	SAS	18
	SAS Notes	18
10.	BMDP	18
	Notes on PROC BMDP IN SAS	18
	P1T Soon to be released	19
11.	MUSIC	19
	Dial-Up Terminal Courtesy A Must	19
	What Is A MUSIC File?	19
	The File Concept: What's The Difference?	20
12.	COMPUTER HUMOR	22
	The End Of OS: Thousands Wept	22
13.	INDEX TO PAST ISSUES	24
14.	WORKSHOP/SEMINAR CHECKLIST	27

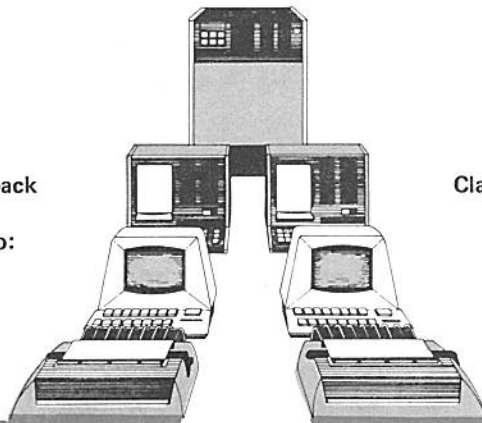
BENCHMARKS Reader/User feedback is encouraged.

Send all letters, suggestions, etc., to:
North Texas State University
The Computing Center
NT Station, Box 13495
Denton, Texas 76203

Claudia Putnam, BENCHMARKS Editor

Richard Harris,
Director of Computer Systems

Thomas Madron, Manager,
Academic Computing Services



SERVICES AVAILABLE TO USERS OF THE NTSU COMPUTING FACILITIES

All people mentioned below may be contacted by calling (817) 788-2324.

Information and Project Numbers - Sandy Franklin or Sue Heffley in the Computing Center Reception Area, ISB 119

Test Scoring and Analysis - Sue Heffley or Sandy Franklin

Newsletter Questions/Contributions/etc. - Claudia Putnam

Statistical/Research Support (provided for graduate students and faculty members) - Bob Brookshire, George Morrow, Claudia Putnam, and Mohamad Salahshour.

Non-Research Student Programming Problems - student consultants from the Computer Science Department found in ISB 134A near dispatch and the user keypunch area. Student consulting provided by the College of Business is available at the BA Computing Access Facility.

JCL and Debugging Problems - Mohamad Salahshour.

Pre-Research Counseling, Including Survey Instrument Design - Bob Brookshire, George Morrow, and Claudia Putnam.

Data Entry - to MUSIC, Keypunch Requests and Questions Regarding Layout of Keypunch Sheets - Betty Grise, ISB 227.

Academic Timesharing Information and/or Problems HP/2000 and AS/5000 MUSIC (McGill University System for Interactive Computing) information and problems, including terminal problems - Mohamad Salahshour.

Administrative Applications - Coy Hoggard.

AS/5000 Computer Hardware/Software/Billing Problems - Sandy Franklin.

JOB Submission and Retrieval - RJE operators.

SPRING COMPUTING HOURS

Computing facilities will be open during the following times throughout the Spring Semester (not applicable to holidays):

Computing Center RJE: 7 AM Monday to Midnight Saturday; Noon to Midnight Sunday.

College of Business RJE: 7:30 AM - Midnight, Monday - Friday; 8 AM - Midnight, Saturday; Noon - Midnight Sunday.

Media Library(GAB): 8:00 AM to 12:00 PM Monday through Thursday;
8:00 AM to 6:00 PM Friday; 11:00 AM to 5:00 PM Saturday; 4:00 PM to 12:00 PM Sunday.

PART III OF USER'S GUIDE AVAILABLE IN THE UNIVERSITY STORE

The January 1982 issue of BENCHMARKS (Volume 3, Number 1) contained an article announcing the fact that the Academic Computing staff is engaged in preparing a users guide to computing facilities at NTSU. One section, Part III : Using OS/MVT is now available in the University Store at a cost of \$.95 (that's right, cents). The rationale behind releasing sections instead of the entire manual, was that they could be combined, by users, into a complete manual, and valuable information would be made available as soon as each section was completed.

TELL US WHAT YOU WANT

During the past school year, Academic Computing Services has offered the following short courses for your benefit.

1. Introduction to MUSIC
2. Introduction to MUSIC/SCRIPT
3. Introduction to SPSS
4. Introduction to SAS
5. Introduction to BMDP
6. Advanced Topics in MUSIC

Please complete the questionnaire at the back of this issue of BENCHMARKS to let us know what your future wants/needs are in this area.

If you have further questions or suggestions, please call us at (817) 788-2324, or send MAIL to the SYSTEM (see the June 1981 BENCHMARKS or enter HELP MAIL while logged on to MUSIC for more information on sending MAIL).

DATA ENTRY INTO MUSIC FILES

The Data Entry section of the Computing Center, located in ISB 227, will now enter data directly into a user's MUSIC files. In order to take advantage of this service, The user is required to have:

1. A valid project number.
2. A MUSIC ID
3. Data in a form acceptable to the data entry personnel.

Should the user desire, a text file can be created for him/her, from a document, using MUSIC/SCRIPT.

It is advisable, since users' passwords will have to be divulged, for them to change their passwords after their files have been entered into their ID's.

800 BPI TAPES REQUIRE SPECIAL SCHEDULING

The processing of 800 bpi tapes, as discussed in the article below, is being phased out here at NTSU. In the interim, it is requested that special arrangements be made to process those jobs requiring the use of the 800 bpi tape drive. This requires that a special handling card be filled out and signed by someone in Academic Computing (either Tom Madron, Claudia Putnam, or their designate). The user will then submit his/her job as a CLASS=N job along with the special handling card. It is important for users to recopy their 800 bpi tapes to a higher density as quickly as possible (see the following article). If it is not feasible to do this, please notify either Tom Madron or Claudia Putnam (788-2324; ISB).

800 BPI PROCESSING COMING TO AN END

The Computing Center currently supports tape processing with four NAS tape drives, three of which read and/or write either 6250 or 1600 bpi tapes. The fourth tape drive reads/writes 1600 or 800 bpi tapes. The long-range plan for tape processing is to do away with the 1600/800 bpi tape drive and replace it with another higher density 1600/6250 drive. To keep from getting caught with an 800 bpi tape and no tape drive to read it, ALL USERS WHO HAVE 800 BPI TAPES SHOULD TAKE STEPS TO RE-COPY THOSE TAPES TO A HIGHER DENSITY (1600 or, preferably, 6250 bpi). This can be accomplished in a variety of ways one of which will be discussed here.

SAS (Statistical Analysis System) has a utility procedure called TAPECOPY which will copy 9 track standard or non-labeled tapes in their entirety or portions there-of onto another tape (also 9 track, standard (SL) or non-labeled (NL)). A complete description of PROC TAPECOPY is available in the SAS USER'S GUIDE: 1979 EDITION pages 419-422. Rental tapes can be obtained from the Computing Center Dispatch personnel (ISB 134). Tapes can also be purchased at the Computing Center Reception Area (ISB 119). All output tapes should be submitted to the computer operator for initialization prior to copying any tapes. Some typical 800 to 6250 bpi TAPECOPY jobs are illustrated below:

SL to SL:

```
//jobname JOB (nnnn-nnnn,:20,1),'user name',CLASS=N
// EXEC SAS
//VOLIN DD UNIT=TAPEDD,DISP=OLD,
// VOL=SER=xxxxxx,LABEL=(,BLP)
//VOLOUT DD UNIT=TAPE9,DISP=(NEW,KEEP),
// VOL=SER=yyyyyy,LABEL=(,BLP)
PROC TAPECOPY;
```

SL to NL:

```
//jobname JOB (nnnn-nnnn,:20,1),'user name',CLASS=N
// EXEC SAS
//VOLIN DD UNIT=TAPEDD,DISP=OLD,
// VOL=SER=xxxxxx,LABEL=(,SL)
//VOLOUT DD UNIT=TAPE9,DISP=(NEW,KEEP),
// VOL=SER=yyyyyy,LABEL=(,NL)
PROC TAPECOPY;
```


SL to SL, selected files:

```
//jobname JOB (nnnn-nnnn,:20,1),'user name',CLASS=N
// EXEC SAS
  /VOLIN DD UNIT=TAPEDD,DISP=OLD,
// VOL=SER=xxxxxx,LABEL=(,BLP)
//VOLOUT DD UNIT=TAPE9,DISP=(NEW,KEEP),
// VOL=SER=yyyyyy,LABEL=(,BLP)
PROC TAPECOPY;
  FILES 1-7 9;
```

It is strongly suggested that you let the tape drive default to the highest density (6250) when copying your tape, and that you create standard label tapes (standard label tapes cannot be created from non-labeled tapes, using PROC TAPECOPY, however - see below). Of course, if there is some specific reason why this should not be done, then create the type of tape that you know will best serve your needs.

The job-setup for copying a non-labeled tape onto a standard labeled tape is as follows:

```
//jobname JOB (nnnn-nnnn,:20,1),'user name',CLASS=N
// EXEC SAS,REGION=200K
//IN DD UNIT=TAPEDD,DISP=(OLD,KEEP),VOL=SER=xxxxxx,
// LABEL=(1,NL),DCB=(RECFM=xx,LRECL=nnnnn,BLKSIZE=nnnnn)
//OUT DD UNIT=TAPE9,DISP=(NEW,KEEP),VOL=SER=xxxxxx,
// LABEL=(1,SL),DSN=xxxx,DCB=(RECFM=xx,LRECL=nnnnn,BLKSIZE=nnnnn)
DATA _NULL_;
  INFILE IN;
FILE OUT;
INPUT;
PUT _INFILE_;
```

The DCB information should be entered in the JCL. If you have more than one file on the non-labeled tape, each file will have to be copied, starting with the first, in a separate run (ex: Run 1 - LABEL=(1,NL) to LABEL=(1,SL); Run 2 - LABEL=(2,NL) to LABEL=(2,SL); etc.). If you do not specify a data set name (DSN) for each file, SAS will manufacture one.

Anyone absolutely positively needing to continue to process 800 bpi tapes should contact either Tom Madron or Claudia Putnam in Academic Computing (788-2324; ISB 119).

SPACE ON ACADEMIC DISK PACKS RUNNING DANGEROUSLY LOW

Space on the academic disk packs (ACAD00, ACAD01, ACAD02) has been running dangerously low. When this happens all users suffer. Disk files cannot be allocated because there is not enough space for them, and jobs bomb because they cannot find enough workspace. In order to remedy this situation, it will be necessary to delete all files containing (a) old project numbers (i.e. those not valid this academic year) (b) non-conventional data set names (i.e. something other than USERn.Dnnn.Pnnn.XXXXX - a complete explanation of the naming convention is given in the March, 1981 issue of BENCHMARKS). The system clean up procedure will be done beginning Friday April 16, 1982. Users were notified by a memo sent out April 1 (not an April Fools joke!).

It is always the user's responsibility to take care of renaming or updating his/her file names (using current project numbers). This

procedure must be followed in order to ensure sufficient storage for all users. Should you have old files which are not used frequently, contact the Computing Center dispatch personnel about renting or purchasing a tape to archive (dump) the files to tape.

It is especially important that files generated by students during a class be purged by the end of each semester. It is the instructor's responsibility to purge those files, although the Computing Center personnel will provide assistance in that effort. Instructors should provide the Computing Center with lists of files generated by any classes for which they are responsible. Disk space is not unlimited and what is improperly used by one person is denied to another.

Please always make sure that your files have conventional names with correct project numbers. If you have a file which does not correspond with our naming system, use our new simple renaming procedure to rename your file (DSN) as follows:

```
//yourjob JOB (nnnn-nnnn),yourname
// EXEC IEHPRGM
  RENAME DSN=old.data.set.name,VOL=SYSDA=volumename,      *
  -                               NEWNAME=USERn.Dxxxx.Pyyyy.newname -
  -                               -                               -
  -                               -                               -
(Col 2)          (Col 16)                                     (Col 72)
```

If you need to delete some data sets simply use the IEFBR14 Utility:

```
//yourjob JOB (nnnn-nnnn),yourname
// EXEC PGM=IEFBR14
//DD1 DD DSN=data.set.name,VOL=SER=ACADnn,
//      UNIT=SYSDA,DISP=(OLD,DELETE)
//DD2 ....data set information .....
//DD3 ....data set information .....
//DDn ....data set infromation .....
```

If any of your data sets are catalogued, simply leave off the volume serial (VOL=SER=ACADnn) in the DD statement, and they will be deleted and uncatalogued.

If you have any difficulties in remembering the names of your files, or if you have any questions please call Mohamad Salahshour (788-2324) or send MAIL to the SYSTEM while logged on to MUSIC.

It is not a ~~bad~~ ^{NEW} idea, when allocating new data sets on disk to use the disposition ~~OLD,KEEP,DELETE~~ ^{NEW} (DISP=~~OLD,KEEP,DELETE~~). This way, under most conditions (there are exceptions), if your job does not run successfully the data set allocated in the JCL will be deleted. This will serve a two-fold purpose: (1)It will keep you from having to run a separate procedure to delete the data set so that you won't get a "Duplicate Name On Volume" message and (2)It will ensure that empty data sets are not propagated throughout the disk packs because you changed your data set's name and neglected to delete the empty one (a very common occurrence!).

USING THE COMPUTER FOR RESEARCH: PART II

By Bob Brookshire

This article is the second in a series of articles on computing and data analysis. The first appeared in the February 1982 issue of BENCHMARKS, Volume 3 Number 2.

* * * * *

The research act is not a random occurrence. It is, or should be, the product of careful planning and preparation. Indeed, the planning and preparation of research is so important that it is recognized in many disciplines as a distinct subfield: research design. Volumes have been written on this topic, some of which are cited at the end of this article. We cannot begin to cover this subject in any detail in BENCHMARKS, but will instead offer a brief outline or review of the major principles of research design, and some of the statistical considerations that should be kept in mind in the planning stage of research.

The Logic of Research Design

Explanation and prediction are probably the two major goals of scientific research. Most explanations and predictions take the form of logical deductive, or more frequently, inductive arguments. The premises of these "arguments" contain generalizations based on known scientific laws or observed statistical relationships, and the conditions of the event for which the explanation or prediction is desired. The conclusion, the event or condition to be explained, follows logically from these premises, either certainly or with a known probability. This structure should be kept in mind throughout the stages of the research process, from the formulation of the research problem, through the data collection and analysis, to the presentation of the results.

If logical argument is the structure on which research is built, problem formulation is the foundation. Scientific research problems may be empirical, analytic or normative in nature, but regardless of their type, they share the characteristics of explicitness, clarity, originality, testability, theoretical significance and relevance. Some researchers find it immensely difficult to formulate research problems which meet all these criteria, but the rewards for following these standards are realized in the ease with which the research is conducted and the "publishability" of the results.

The result of the problem formation stage of research is a scientific hypothesis. We might define a hypothesis as an empirically testable statement derived from a theory. Subsumed in this definition are the requirements that a hypothesis be conceptually clear, operationally defined and related to available research techniques. Researchers using statistics most often formulate a specific type of hypothesis called the "null hypothesis," which states that the occurrence of a variable in a population is zero, or that there is no relationship between two variables. Statistical tests are generally constructed to attempt to reject this research hypothesis. Obviously, the type of hypothesis developed by a researcher determines to a great extent the methods used in the investigation. Careful hypothesis formation is crucial to the successful completion of a research project.

The concept of "theory" brings us back again to the first part of this section, explanation and prediction. If we think of a theory as a set of hypotheses that are deductively connected, then the relationship between explanation, hypotheses and theory becomes clear. Explanation is accomplished through the construction of a theory by the testing of a set of deductively related hypotheses. The foundation of the research act, problem formulation, supports the building blocks, the hypotheses. The logical relationships between the hypotheses yield a theory, the structure. The whole edifice, then, is the explanation, the goal of scientific research.

Some of the greatest of modern philosophers have written on the logic of scientific research. A readable introduction to this topic is provided in the book by McGaw and Watson, listed in the references at the end of this article. The interested reader can also explore the works by Popper and Hempel.

Statistical Considerations in Research Design

Certain research designs by their nature require certain statistical procedures; likewise, certain statistical procedures assume the existence of specific research designs. At other times, there may be a variety of statistical techniques appropriate to a given investigation, just as some techniques are applicable to a broad range of research designs. This section will offer some pointers to aid the researcher in selecting those techniques that are suitable for his research problem. Before going into any detail, however, some general comments are in order.

As mentioned in the beginning of this article, most computer programs will analyze just about any sort of data. It is your responsibility to insure that the analysis is the appropriate one. It is also your responsibility to interpret the results of the analysis correctly, which means that you should have a reasonable understanding of the methods you have chosen. A good rule of thumb might be not to use a statistical procedure unless you can explain it, at least in applied terms, to somebody else. If you are called upon to defend your research, you'll probably have to do this anyway.

* * * * *

There are many types of research designs. Some are used frequently in one discipline, and not at all, or rarely, in other fields. Others are of such broad applicability that one finds examples in every area of research. Generally, research designs have been classified into two types: experimental, and quasi-experimental. Within these two categories are many variations, which depend mainly on the research setting and the amount of control that the investigator can introduce into the study. Certain statistical procedures have been developed specifically for certain designs, and are difficult, if not impossible, to apply to others. For example, the analysis of variance is particularly suitable to experimental designs, and is not appropriate in many quasi-experimental situations. Some regression techniques assume the existence of large sample sizes, and are therefore not often useful in experimental designs.

Aside from statistical considerations, however, the type of inferences that can be made about the relationships between variables in populations vary substantially with the type of research design. Therefore, the selection of a particular experimental or

quasi-experimental design not only has statistical, but also theoretical implications. You should investigate carefully the assumptions that lie behind the design you have chosen, and select the statistical procedures and theoretical arguments which are suited to them. A complete discussion of types of research designs is provided in the book by Campbell and Stanley, cited below. Good introductory works are Nachmias and Nachmias, (Ch. 4,5) and McGaw and Watson (Ch. 14), also in this list of references.

* * * * *

The most basic statistical, and theoretical, consideration in research design is the identification of the units and levels of analysis. The unit of analysis is simply the entity on which the research observations are made. Most often, the unit of analysis is the individual--an experimental animal or human subject, a survey respondent, a plant in a control or experimental group. At other times, the unit of analysis might be a collection of individuals--a classroom of students, a corporation, a city, state, or country, a jar of fruit flies. The level of analysis is defined as the level at which the analysis is conducted. This might also be at the level of the individual or at the level of a group of individuals. Much research is carried out in which both the unit and level of analysis are the same. Investigators take measurements of individuals, and then from these measurements analyze the characteristics of individuals, or take measurements of aggregations of individuals, and generalize about the characteristics of these aggregations.

Sometimes, however, a researcher wishes to take measurements at one level, and analyze at another. An example would be a sociologist who has data on ethnicity and literacy rates for states, and wishes to look at the relationship between ethnicity and literacy in individuals. Another example would be a political scientist who interviews samples of individuals in different countries regarding their attitudes toward foreign policy, and wishes to compare the foreign policy attitudes of nations. This type of cross-level research must be conducted with extreme care. Inferences about relationships at one level of analysis which are drawn from units which are measured at another are particularly prone to fallacy. This is not to say that this type of analysis is impossible, but only that it is not by any means straightforward. In some cases, special statistical models can be employed which can correct for the attenuation of relationships that occur when units and levels of analysis differ. In other cases, it may be that the desired level of analysis cannot be conducted at all with the available units, or if it is conducted, the results will be merely suggestive or exploratory, rather than definitive. Finally, it may be that a researcher can fruitfully recast his research design so that the units and levels of analysis are the same. This can sometimes yield new and startling theoretical insights. The point of this discussion is to alert you to this potential problem, so that you can take steps to deal with it if it occurs in your research. A good introduction to this topic is provided in the works by Robinson, Goodman, and Przeworski and Teune, cited below.

* * * * *

A more common problem in the selection of statistical techniques arises from the levels of measurement employed in making research observations. If measurement is the assignment of numbers to characteristics, levels of measurement refer to the degrees of

discrimination between similar characteristics that this assignment affords.

Some characteristics can only be discriminated at gross levels. State of birth, sex, and political party registration are some examples of these characteristics. We cannot say that a Democrat has more party registration than a Republican, or that a female has more sex than a male. If we coded the values of sex as 1 for females and 0 for males, that will not indicate that males have no sex. The assignment of numbers to the categories is arbitrary, and any numbers will do the job. We cannot (logically) perform mathematical operations on the numbers that represent these categories; there is no mean or standard deviation for sex or state of birth. It is common to say that these variables are measured at the nominal level, since any numbers of the categories represent the names of the categories, and not any mathematical properties associated with them.

Other variables have categories which can be rank ordered, and so are measured at the ordinal level. Many variables in the social sciences, and a few in the physical sciences, are measured at this level. Some examples are religiousity, conservatism, or some scales of hardness; we can identify units of analysis as more or less conservative, more or less religious, or softer or harder, than other units. With this level of measurement, however, we cannot equate the distances between any two points on the scale, or tell the degree to which any one unit is more conservative, religious or harder than another. We might know that a diamond is harder than glass, but we do not know how much harder it is. One survey respondent may be more religious than another, but we would be unable to say whether he was twice as religious, ten times as religious, or 6.458 times as religious as any other survey respondent. We might assign number values to the categories of hardness, with, for example, glass getting a value of 4 and diamond getting a value of 8. Perhaps William F. Buckley would receive a score of 100 on our conservatism scale, while Edward M. Kennedy received a 1. The relative size of the numbers indicates that one unit of analysis has more of the variable in question than the other, but the magnitude of the numbers is essentially meaningless. We could just as well have given the diamond a 100 and the glass a 1, Buckley the 8 and Kennedy the 4. Thus it is, strictly speaking, inappropriate to apply mathematical operations to variables measured at the ordinal level, just as it is inappropriate for nominal level variables. If we can't say that a person is twice as conservative as another person, how can we multiply the number for his assigned category by two and expect to achieve anything meaningful by the result?

Most variables in the physical sciences, and a few in the social sciences, are measured at the interval or ratio levels. The numbers assigned to categories at these levels of measurement show the degree to which the property associated with the variable is present in the unit of analysis. The intervals between the categories are of equal length, and so mathematical operations can be performed using these measures. One hundred degrees is twice as hot as fifty degrees, four times as hot as twenty-five degrees, and so on. A sample of research subjects can have a mean age, temperature, height, weight, etc. Ratio level measurement differs from interval level measurement in that ratio measures have a true zero point that represents the complete absence of the characteristic being measured. An object can, at least theoretically, have a height or weight of zero, thus height and weight are measured at the ratio level. The temperature scales of Fahrenheit and Celsius are not ratio measurements, in that zero on these scales

does not indicate the absence of temperature. These scales, then, are measured at the interval level.

The importance of interval and ratio levels of measurement lies in their ability to be manipulated mathematically, therefore being suitable for the application of statistical procedures. If you are planning to use statistical methods to analyze your data, you should insure, at the stage of the research design, that the levels of measurement planned for the variables are suitable for the methods to be used. Discussions of levels of measurement are standard in many research design texts. Further information on these topics, and additional references, can be obtained in the works by Nachmias and Nachmias, McGaw and Watson, or Babbie, cited below. The original exposition of these concepts is in the article by Stevens.

* * * * *

The object of most research is to make generalizations based on specific cases. Chemists study reactions between substances in the laboratory setting, and generalize their findings to substances in all settings, not just the laboratory. Agronomists examine crops grown at experimental stations, and generalize their observations to crops grown on commercial farms. Psychologists measure the responses of experimental groups and generalize to the responses of people in everyday interactions. Sociologists and political scientists study samples of members of social and political groups and generalize about the distributions of behaviors and attitudes in entire populations.

When the object of research is inference from a small number of cases, on which analysis is conducted, to a large number of cases, or a population, on which it is difficult or impossible to conduct analysis, the selection of the sample of cases as research subjects is very important. The purpose of the variety of sampling techniques that have been developed by statisticians is twofold: first, bias in the sample is at all costs to be avoided; second, the sample should be representative of the population from which it is drawn. Simple random sampling, the most familiar sampling technique, assures, at least theoretically, that the first of these goals is achieved. If research subjects are chosen truly at random, then no bias is intentionally introduced--any bias in the sample has occurred merely by chance, and is not the result of anything the investigator has done. Most inferential statistical procedures assume that simple random sampling has been used to draw the subjects on which estimates are based. Other sampling techniques, like multi-stage sampling, cluster sampling, or stratified sampling, attempt to combine knowledge that the researcher has, a priori, about the nature of the population to which inferences are to be made, with the theoretical utility of randomness. Consequently, making inferences on the basis of samples of this type is a more complicated process, but yields estimates of population characteristics that are more precise than simple random sampling. You should carefully choose the sampling techniques for your projects, and be prepared to make any necessary adjustments to your estimates of population characteristics that these techniques require. If you are interested in further information about sampling techniques, you should consult the books by Cochran, Kish or Warwick and Lininger, cited below.

Statistical Analysis

We have already discussed the formulation of hypotheses, and how these hypotheses are related to the goals of research, explanation and

prediction. After the hypotheses have been formulated, and the general strategies of research are being devised, methods must be selected to test the hypotheses. This section will introduce the general procedure for testing hypotheses, and show the relationships between various levels of measurement and some of the statistical procedures suitable for making these tests at these levels of measurement. Naturally, it is impossible in a work of this scope to describe adequately a wide variety of statistical methods. Rather, this presentation is designed to show, for those who are having trouble selecting a procedure, the general directions they can consider. Different disciplines vary widely in the types of procedures used in the "mainstreams" of their research. For these reasons, we have tried to keep this discussion at a level of generality that is applicable to a broad range of research interests.

A research hypothesis is an assertion that some condition is true, for example that there is a relationship between two variables, or that two groups differ with regard to some characteristic. Statistical tests can rarely, if ever, demonstrate the truth of a research hypothesis. They can show, however, with a known probability, that a null hypothesis, a statement that there is no relationship between two variables, or that two groups do not differ, is false. Thus, the rejection of a null hypothesis offers some evidence in support of the research hypothesis, although it cannot prove it, and it eliminates the alternative hypothesis that there is no relationship, or that an observed condition is due to chance.

Note, however, the phrase "with a known probability" in the statement above. Tests of null hypotheses can never completely rule out the possibility that the null hypothesis is true, but can only do so with a certain level of probability, known as the significance level of the statistical test. This significance level is chosen by the researcher before the statistical test is made, and is customarily .05 (.01 in some disciplines). This means that the researcher will reject the null hypothesis if it is false with a probability of 95, but will not reject it if the probability is less. He is willing to take a 5 risk that the null hypothesis is true.

What happens if the investigator loses, and rejects a true null hypothesis? He has committed what is known as a Type I error, and his research conclusions are wrong. Perhaps he would not have made this error had he selected a different significance level, .01 for instance. With this significance level, he would reject the null hypothesis only if it is false with a probability of 99 or greater. By selecting this more stringent significance level, however, the researcher has increased the probability of making another kind of error, called a Type II error. This is the fallacy of failing to reject a null hypothesis that is false. These two errors are, obviously, inversely related. The more researchers try to reduce the occurrence of Type I errors, the more they increase the possibility of a Type II error, and vice versa. In the main, though, scientists are primarily concerned with avoiding making Type I errors, so they keep the criteria for rejecting null hypotheses very strict. They would rather avoid punishing an innocent null hypothesis, and so let a few guilty ones get by.

How does a researcher select the statistical methods used to test his hypotheses? One important consideration, as mentioned above, is the level of measurement of his variables. Another important factor is the number of variables which are included in the analysis. Other criteria include assumptions about the theoretical distribution of the variables in the population under consideration, and the power (the ability to

detect significant differences) of the statistical procedures. The following chart shows some statistical procedures that are often used to test hypotheses or describe relationships with different numbers of variables at different levels of measurement. Not all statistical procedures are included, and many disciplines make practical use of procedures that, strictly speaking, are not always appropriate for certain levels of measurement. The final choice of a procedure depends on the needs of the investigation and the standards of the discipline, and so might vary in different situations which are seemingly similar. This list, therefore, is mainly suggestive, rather than exhaustive or definitive.

Statistical Procedures for Various Levels of Measurement and Numbers of Variables

1. Univariate (single variable) procedures

Level of Measurement	Examples of Procedures
Nominal	tests of proportions, ratios
Ordinal	tests of medians, quartiles
Interval & Ratio	tests of means, variances, etc.

2. Bivariate (two variable) procedures

Level of Measurement (variable 1)	Level of Measurement (variable 2)	Examples of Procedures
Dichotomy	Dichotomy	Fisher's exact test, odds ratio
	Nominal	Chi Square
	Ordinal	Runs test, Mann-Whitney U test
	Interval & Ratio	T-test, logistic regression
Nominal	Nominal	Chi Square
	Ordinal	Friedman's test, Kruskal-Wallis
	Interval & Ratio	Analysis of Variance
Ordinal	Ordinal	Spearman's Rho, gamma
Interval & Ratio	Interval & Ratio	Correlation, Regression

3. Multivariate (three or more variables) procedures

A. One variable considered dependent

Level of Measurement of Dependent Variable	Level of Measurement of Predictors	Examples of Procedures
Nominal	Nominal	Grizzle-Starmer-Koch discriminant function
	Interval & Ratio	
Interval & Ratio	Nominal	MANOVA, regression
	Interval & Ratio	MANCOVA, regression

B. No variable considered dependent

Level of Measurement	Examples of Procedures
Nominal	log-linear modelling
Ordinal	multidimensional scaling
Interval & Ratio	factor analysis

This concludes our discussion of research design and the statistical considerations which attend it. We hope that it will prove helpful to those who are about to begin their first research projects, as a guide to further investigation of this important subject. We hope, too, that it may be useful to those who are seasoned investigators, as a sort of refresher course, and, possibly, an introduction to new methods of analysis. The reference list will direct interested readers to further information on both research design and statistical methods.

References

- Babbie, Earl R. The Practice of Social Research. Belmont, Cal.: Wadsworth, 1975.
- Blalock, Hubert M. Social Statistics. New York: McGraw-Hill, 1979.
- Campbell, Donald T. and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1963.
- Cochran, William G. Sampling Techniques. New York: John Wiley & Sons, 1977.
- Goodman, L. A. "Ecological Regressions and Behavior of Individuals." American Sociological Review 18 (Dec. 1953): 663-664
- Hempel, Carl. The Philosophy of Natural Science. Englewood Cliffs, N.J.: Prentice-Hall, 1966.
- Kish, Leslie. Survey Sampling. New York: John Wiley & Sons, 1965.
- Kmenta, Jan. Elements of Econometrics. New York: Macmillan, 1971.
- McGaw, Dickinson and Watson, George. Political and Social Inquiry. New York: John Wiley & Sons, 1976.
- Nachmias, David and Nachmias, Chava. Research Methods in the Social Sciences. New York: St. Martin's Press, 1981.
- Popper, Karl. The Logic of Scientific Discovery. New York: Harper & Row, 1959.
- Przeworski, Adam and Teune, Henry. The Logic of Comparative Social Inquiry. New York: John Wiley & Sons, 1970.
- Robinson, William. "Ecological Correlations and the Behavior of Individuals." American Sociological Review 15 (June 1950): 351-357.
- Stevens, S. S. "On the Theory of Scales of Measurement." Science 103 (1946): 677-680

Sullivan, John L., general editor. Sage University Paper series on Quantitative Applications in the Social Sciences. (25 papers) Beverly Hills: Sage Publications.

Warwick, Donald P. and Lininger, Charles A. The Sample Survey: Theory and Practice. New York: McGraw-Hill, 1975.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1971.

Wonnacott, Ronald J. and Wonnacott, Thomas H. Econometrics. New York: John Wiley & Sons, 1970.

```

* * * * *
*                               *
*               SPSS           *
*                               *
* * * * *

```

RELEASE 9.1 INSTALLED

Release 9.1 of SPSS has been installed and is now the procedure accessed when you // EXEC SPSS. This is simply an updated version of Release 9. The following errors, some reported here for the first time have been CORRECTED in this release. These are being printed in case someone "out there" got one of these errors previously and did not realize it was an error.

BOX-JENKINS:

1. If Q=1 and ARLAG=12, the value printed in the label MA LAG was wrong. The number 12 was printed when it should have read 1.
2. Supplying an invalid variable name in the VAR= syntax could have produced a system abend after the SPSS-generated diagnostic messages.

INPUT FORMAT :

1. The input format correspondence table columns for A-type variables did not align with other types.

MANOVA :

1. When BOXPLOTS were selected, sometimes the analysis would abend with a system error. Other times, the detrended normal plot might drop out-of-range values and on the normal plots the out-of-range symbol might be put in the wrong position.
2. When MANOVA terminated due to insufficient workspace, meaningless SPSS error messages would be produced and the run would have an abnormal completion.
3. Sometimes there would be an alignment error due to an incorrect typing of arguments, which caused an alignment error message.
4. The notification of insufficient workspace for a MANOVA run was confusing. It is supposed to be more understandable now.

MULT RESPONSE :

1. PAIRED - Requesting percentages on a table with empty columns caused divide checks.

NEW REGRESSION :

1. A number of minor changes have been made to this procedure, including rewording warning messages and error messages, tuning

the code to make it less expensive to execute, and including new numerical checks for added dependability.

ONEWAY :

1. Specifying a range greater than 50 on the independent variable caused a wealth of problems. Now a check is performed on the syntax and a warning notice is printed and the analysis is terminated if the range is too large.

REPORT :

1. Under very unusual and poorly understood circumstances, the IBM FORTRAN H compiler used by SPSS Inc. would generate bad object code for routine RPSTAT with optimization level 2. All the possible symptoms are not known, but they included divide checks, overflows, and other FORTRAN errors. Compiling this routine with OPT=1 circumvents the problem, which was done in in Release 9.1.

KNOWN ERRORS IN SPSS RELEASE 9.1

The following list contains all of the known errors, so far, in Release 9.1 of SPSS. Any errors that were documented for a previous release and are not listed here have been fixed.

Reading Data :

If there are multiple records per case and the last case is short a record, there is no error message, but the last case is dropped.

BREAKDOWN :

1. Option 2 (Exclude cases with missing values on the dependent variable) does not work for CROSSBREAK. Cases with missing value on the dependent variable are included in the calculation of statistics.
2. In integer mode - if a huge problem is specified, the workspace calculation overflows and generates an IBM system error OC4.
3. For CROSSBREAK, there is a maximum of 200 non-empty rows and 200 non-empty columns allowed. If these maxima are exceeded, a non-fatal message is printed and processing continues.

CANCORR :

1. For moderately large problems, there may be an incorrect calculation of canonical variate scores and occasional overflows.

COMPUTE and IF :

1. An expression of the form $Y - 2 ** Z$ is parsed incorrectly as $Y + (-2) ** Z$. The result is correct if Z is odd, but incorrect if Z is even. Only expressions exactly like this are affected: those where a constant to a power are subtracted from a variable or an expression. Use parentheses to force correct parsing: $Y - (2 ** Z)$.
2. Just COMPUTE - An expression of the form $A/(200*.*B)$ does not produce an error.

CROSSTABS :

1. Option 7 produces incorrect page numbering.
2. General mode - if the last table requested is empty, no message is printed.

DISCRIMINANT :

1. Occasionally two different scales are used for the territorial map and the scatter plots.

DOCUMENT :

1. If two files with documents are merged, the document text from the second file seems to disappear.

EDIT :

1. No warning is issued about too many variables.

MANOVA :

1. Plotting routines may produce extra headers. Setting PAGESIZE to something larger than the default 55 and less than or equal to the actual number of lines that fit on your printer paper seems to help.

MULT RESPONSE :

1. All reported frequencies have minima of 1.
2. Contains no check for alphanumeric variables.

NEW REGRESSION :

1. The following are null operations in NEW REGRESSION, but no warning message is produced. Instead, the request is ignored : Residuals keyword ID specified without a variable name; Residuals subcommand SCATTERPLOT with no pairs of variables specified; Residuals subcommand SAVE with no keywords specified.

NPAR TESTS :

1. For all tests - if all of the variables used have ranges of missing values (no variable has just one missing value), Option 1 may be forced on by mistake.

PRINT BACK :

1. A PRINT BACK command immediately preceding a LIST CASES causes a spurious LIST CASES error.

RELIABILITY :

1. Friedman's test (Option 15) cannot be computed by the default computing method because that code skips scales with zero variance, which occur whenever the data are ranks. When Option 15 is selected, Option 14 should be selected as well, to force the alternative computing method.
2. Method 2 (covariance matrix method) attempts division by zero if Statistic 9 (item-total statistics) is requested and a variable has no variance or the scale has no variance with the item deleted.

REPORT :

1. If a string variable is used as a break and the break column width is forced to be less than the string width, the BREAK breaks on every case prematurely, even if the string value has not changed.

SAVE ARCHIVE or LIST ARCHINFO :

1. Will fail if DOCUMENT precedes GET ARCHIVE in the run.

SEED :

1. The seed produced by the PRINT option cannot be reproduced.

SORT CASES :

1. Will attempt to sort after an error in file definition, generating error 1778 on FT02.

VALUE LABELS :

1. If a variable name is omitted on the VALUE LABELS card, the value labels of the missing variable destroy the value labels of the previously defined variable instead of generating a syntax error.

WRITE FILEINFO and LIST FILEINFO

1. Both commands produce incorrect N OF CASES values if they are the first procedure and if N OF CASES is UNKNOWN.

It is important that any suspected errors in SPSS (incorrect calculations, "funny" output, etc.) be reported to the SPSS Coordinator here at the Computing Center (Claudia Putnam), since that is one of the primary ways that SPSS has of identifying possible bugs in their procedures.

1. PROC FREQ will only compute statistics for two-way tables. The manual is not clear on this subject.
2. Homogeneity of variance statistics are abundant in SPSS, however in SAS, this is not the case. For two groups, PROC TTEST computes an F (folded) statistic to test for the equality of variances. For two or more groups, PROC TABLES (described in the SAS Institute Technical Report 1-120x, February 1, 1981) will compute a Bartlett's test for a control group and up to nineteen treated groups.
3. Normally, in SAS, when the data are in-stream (following the data definition statements <INPUT, IF statements, etc.>), the SAS statement CARDS; preceeds the data. There may be times, however, when it is desirable to use an INFILE statement along with the CARDS; statement. This would be useful if you wanted to take advantage of some of the functions of the infile statement, such as the LINESIZE, FIRSTOBS, OBS, START, or MISSEVER options. Your job setup would look something like this:
DATA; INFILE CARDS MISSEVER; INPUT VAR1-VAR10; CARDS; etc.....

SAS PROC BMDP can be used to convert a SAS file to a BMDP file and/or execute a BMDP program. Ordinarily, SAS PROC BMDP makes only the SAS numeric variables available to BMDP. SAS also has character variables. One or two of the SAS character variables can be used as BMDP case label variables. To do so, you will need the LABEL and LABEL2 options described on page 138 of the SAS USER'S GUIDE: 1979 EDITION.

In addition, in the current version of SAS, you must include a SAS VAR statement with the names of all the variables that you want to be available to BMDP. An error will be generated if you specify LABEL and do not have the variable in question included in the SAS VAR statement.

P1T SOON TO BE RELEASED

There is good news for our users who have been waiting for our univariate and bivariate spectral analysis program. P1T will be ready for release in the first quarter of 1982. It has already been included in our latest documentation, the 1981 BMDP Manual and the new 1982 User's Digest, so that the information will be readily available to you when P1T is released.

```

* * * * *
*
*           M U S I C
*
* * * * *

```

DIAL-UP TERMINAL COURTESY A MUST

The closer to the end of the semester we get, the busier and busier the system becomes. This causes two things to happen to MUSIC: (1) response time gets slower and (2) all the ports for dial-up terminals become busy. The latter problem can be alleviated to some extent if users will not leave dial-up terminals idle and connected to MUSIC for longer than 5 minutes at a time. This will ensure that others may dial in and get connected to MUSIC. When all the dial-up ports are in use, you will either get a busy signal or the telephone will continue to ring, but MUSIC will not answer it.

WHAT IS A MUSIC FILE?

The MUSIC interactive operating system can be thought of as a house with many rooms, one of which is full of filing cabinets (ID's). Each MUSIC user gets a filing cabinet assigned to him/her that will hold a specified volume of material. The user then proceeds to fill the filing cabinet with file folders (MUSIC files). The user creates files by entering data, text, and/or programs into the terminal while logged on to his/her ID, and saves these files by labeling them. Any number of files may be contained within the filing cabinet, as long as they do not exceed the volume the cabinet can hold. The primary interaction between most users and MUSIC, then, consists of creating and manipulating files.

If one were doing batch processing using cards, one could have a series of card decks, each performing a specific function, stored in separate file folders. Text could also be stored in the folders in the form of letters, term papers, etc. The file folders...

identifiable by some label that the owner had placed on the outside of the folder. The label on the folder may or may not have anything to do with the contents of the folder, depending on the whim of the owner. The contents of the folders could change without the labels changing since it would be up to the owner to keep the contents and the labels to date.

In the tangible world, the owner of the filing cabinet would simply manipulate the files by hand. In the abstract world of interactive computing, however, the owner of the filing cabinet cannot actually touch the files. Instead, (s)he must instruct intermediaries (editor programs) to access and manipulate his/her files and this must be done a specific way (through a terminal) in a prescribed manner (by issuing various commands). It is as if the "house of MUSIC" is in a protective bubble, with supplicants attending it. The user must go through the supplicants to get any computing done.

MUSIC commands are issued by the user so that (s)he can gain access to the filing cabinet and its files. The user can then manipulate these files by entering Context Editor commands. The contents of files can be altered, copied to other files, merged with other files, etc. by issuing context editor commands. Programs contained within MUSIC files can be run by issuing MUSIC commands.

As was stated earlier, an unlimited number of files can be contained within the filing cabinet, as long as they do not exceed the pre-specified volume limit of the cabinet. When this volume is exceeded, the user will receive a message stating, in effect, that there is no more room in the filing cabinet. Should this happen the user has two choices:

- (1) Clean out the filing cabinet by:
 - (a) throwing away unwanted files (PURGE them) and/or
 - (b) removing seldom used files and storing them someplace else - perhaps by using the #EXCHANGE utility.
- (2) Appealing to the MUSIC Coordinator for more space - usually not very successful.

THE FILE CONCEPT: WHAT'S THE DIFFERENCE?

If a program contained within a file is going to run native under MUSIC, the output will automatically appear on the user's terminal or be sent to a user-designated file. If, on the other hand, a program needs to run under another operating system (most likely OS/MVT), MUSIC will pretend it is a card reader, and send the program to OS/MVT. This is where the file concept can become quite confusing.

If, for example, a user has an SPSS program stored in a MUSIC file, the only thing MUSIC can do with that file is store it, PURGE it or submit it to OS/MVT. Using SPSS through MUSIC, it is possible to have four different types of files:

- (1) The MUSIC file that contains the card images of the SPSS job plus a line to tell MUSIC to submit the job to OS/MVT for processing (/INCLUDE OSJE) and a line that serves the same function as the "pink card" placed at the end of a card deck (SYSTEM='OS',TYPE='faculty,student,or staff' depending on your

classification, and optionally ,RETURN to tell MUSIC to return the output of your program to OSJR). <This file is named by the user and can be called anything the user desires as long as it fits the naming conventions established in MUSIC>

- (2) An SPSS system file. This file is either being created or was created in a previous run. There are two levels to this file:
 - (a) The internal label of the system file (found on the SAVE FILE and GET FILE lines). This internal label of the SPSS system file, like the MUSIC file, is given a name that is, hopefully, meaningful to the owner of the file and fits the SPSS file naming conventions. THERE IS NO INHERENT REASON FOR THE MUSIC FILE AND THE SPSS SYSTEM FILE INTERNAL LABEL TO HAVE THE SAME NAME - but they could, if the user wanted them to.
 - (b) The external label of the disk data set that houses the SPSS system file. This is the way OS/MVT knows where to find your SPSS system file, and it is the name (DSN) that appears on the //FT03F001 DD or //FT04F001 DD line in your card-image file. When you want to access a previously created SPSS system file, for example, you would have a line similar to the following in your card-image file, along with a GET FILE statement:

```
//FT03F001 DD DSN=D9999.P9999.my.spss.file1,UNIT=SYSDA,  
// VOL=SER=ACADO2,DISP=SHR
```

This label is also created by the user, in the DSN= portion of the line, at the time that you are saving your SPSS system file, and it can be any name you want as long as it follows the naming conventions established by the NTSU Computing Center (see BENCHMARKS March 1981, Vol 2, No. 2 for a complete discussion of the naming convention for disk data sets). Again, THERE IS NO INHERENT REASON FOR THE MUSIC FILE, THE SPSS SYSTEM FILE INTERNAL LABEL AND THE SPSS SYSTEM FILE EXTERNAL LABEL (DSN) TO HAVE THE SAME NAMES - but they could, if the user wanted them to.
- (3) A "raw-data file" that is contained on either disk or tape. This is the file that is read with an //FT08F001 DD line, and it takes the form:

```
//FT08F001 DD DSN=USERn.Dnnnn.Pnnnn.name,UNIT=SYSDA,  
// VOL=SER=ADACOn,DISP=(OLD,KEEP)
```

for disk data sets, and:

```
//FT08F001 DD DSN=name,VOL=SER=tapename,  
// LABEL=(file#,sl or nl),UNIT=TAPE9
```

for tape data sets. Note that the disk data set is named according to the same convention as that for the external label of the SPSS system file. This is because they are both, through job control language (JCL), communicating with the OS/MVT operating system, and the naming conventions for disk data sets on OS/MVT are the same, no matter what the contents of the data set may be. Once again, THERE IS NO INHERENT REASON THAT THE MUSIC FILE AND THE RAW DATA FILE (DSN) SHOULD HAVE THE SAME NAMES - but they could, if the user wanted them to.
- (4) A MUSIC file containing card images of the user's data. This file would be treated as card input to SPSS and it would be called into the SPSS card-image control statement file by inserting a /INCLUDE filename line directly after the

READ INPUT DATA line of the SPSS job. THIS MUSIC FILE AND THE FILE CONTAINING THE CARD-IMAGES OF THE SPSS JOB CANNOT HAVE THE SAME NAME, SINCE THEY ARE BOTH MUSIC FILES.

Thus, we can see that we have four distinct files: (1) a MUSIC file that contains the card images of an SPSS job that can be edited and submitted to OS/MVT to be processed. (2) An SPSS system file with two levels that was created when a (a)SAVE FILE line was included in a job that also had a (b)//FT04F001 DD line in it, (3) a raw data file, on either disk or tape that is read with a //FT08F001 DD line, and (4) A MUSIC file containing card-image data that is called into the MUSIC file containing SPSS card-image control statements by inserting a /INCLUDE filename line directly after the READ INPUT DATA line of the SPSS job.

```

* * * * *
*
*      COMPUTER HUMOR
*
* * * * *

```

THE END OF OS: THOUSANDS WEPT*

In the first few months of the year that System/360 Operating System came to a full stop, all signs appeared normal and there was no indication of an impending disaster. The SDD Manager of programming systems stated at the spring SHARE meeting that the J Level of FORTRAN V would definitely be implemented, and would at least equal the speed of the E Level FORTRAN V subset, provided it was run on a Model 75 or greater. There was no truth, he asserted, to the rumor that IBM was dropping FORTRAN in favor of PL/3. Option 89, of MVC (multiprogramming with a variable number of cpu's), which had been released in System Release 101.8, was hailed by a large number of users as the ultimate in operating systems. Representatives of a major government agency, which had been running a Model 91 with 8 million bytes using a modified BPS supervisor, lodged a mild protest but were shouted down by the majority.

On April 1, an announcement by the Management Information Department of DPD caused quite a stir. Their Management Action Optimization (MAO) program would be written using the new Linear Interpretation Nucleus (LIN), part of DOS extended. This occurred, it was rumored, in spite of persistent efforts by the Marketing Verification Department (MVD) to persuade them to use OS. This department, sometimes called the Red Guard or simply "the butchers," is charged with "purification" of Type II programming standards.

There were indications, however, that something was in the air. The OS Internals Workshop was extended from 13 weeks to 26 weeks. A resident psychiatrist was installed to try and cut down on nervous breakdowns, defections, and AWOL's. A blue letter advised salesmen that "throughput" and "turnaround time" were now verboten in the company's vocabulary. The byword was to be "full utilization of system resources." At all costs customers were to be discouraged from asking, "But when will my job be completed?"

Release 91.0 contained a module of the nucleus that stopped the software clock during system overhead time. Murmurs about the difference between meter time and time-accounted-for led to removal of all meters and a shift from a 176-hour base to 264 hours per month. And

dissatisfaction increased. One large scientific/engineering/commercial customer announced his intention to switch to a competitor, but after two years was unable to do so because he could not discover exactly what his system was doing.

The end finally came in mid-October. System Release 110.7 was distributed, which converted everyone to MPSS (Multiple Priority Scheduling System), which combined the following control program options:

- * Multiprogramming with a variable number of tasks
- * Multijob initiation
- * Multiple priority selection
- * Multiprocessing with a variable number of cpu's

System was accomplished with little difficulty in no more than 504 system hours. Expectantly, customers IPLed and initiated their job streams. And nothing happened.

Nothing.

When it slowly dawned on everyone that nothing was going to happen, now or later, a flood of anguished telephone calls swamped the branch offices. At Poughkeepsie, in turn, all extensions, all 25,000 of them, were busy. Unauthorized vehicles were turned away at the entrance roads. The director of programming systems in Harrison could not be found.

At last a brave man, a customer engineer, fought his way through the crowd around his system and obtained a dump. As he scanned the hex, the horrible truth came home to him. All of core, as far as the eye could see, was filled with control blocks, each containing pointers to other control blocks. DADSM was allocating and suballocating, searching DSCB'S and building new ones. Job management was initiating new jobs, task management was creating tasks and ATTACHing and LINKing, data management was opening data sets, and building WTG tables, DCB's, DEB's, ECB's, and IOB's. It was finding TIOT's from tasks dispatched by task management, which pointed to JFCB's. But no programs were being executed. No data was being read or written or processed. Operating System had taken over all the system resources and was entirely occupied with issuing supervisor calls, saving registers, restoring registers, chaining forwards and backwards and following pointers all over core. Every pointer led to some other pointer. Operating System, after several years of effort by thousands of programmers, had finally become a closed system.

The great dilemma was solved only through the intervention of the chairman of the board, who personally issued a black-boarded blue letter announcing the withdrawal of Operating System. A large bonfire was built in the Poughkeepsie parking lot in which a huge mountain of OS documentation was burned, while the local high school band played a funeral dirge. Users all over the world wearily set to the task of rewriting using BPS Assembler. A new programming system was announced for delivery in two years, to be called Assembler Stacked Support (ASS). And everyone breathed a great sigh of relief and was happy for a time.

*Source unknown

INDEX TO PAST ISSUES

In order to utilize BENCHMARKS to its fullest capacity and avoid redundancies, an index of previous issues containing information considered still pertinent to the NTSU Computing Center is included in each issue.

<u>SUBJECT</u>	<u>VOL/NO</u>	<u>MONTH/YEAR</u>	<u>PG</u>
1. ACADEMIC COMPUTING IN GENERAL			
Exactly What is "The System ?"	3/2	February/82	3
Using the Computer for Research:Part I	3/2	February/82	3
Recommended References for Research	3/2	February/82	3
Directions	3/1	January/82	3
Expanding Our Resources:EDUNET	3/1	January/82	5
Off-Campus Courses of Interest	3/1	January/82	15
Instructional Slide/Video			
Presentation Now Available	2/7	Nov/Dec/81	4
Know Your Representative	2/6	October/81	5
Cooking With Canned Programs : A			
Guide To Instruction In The Stat Paks	2/6	October/81	14
SAS Proceedings Available for			
Examination	2/4	June/81	13
Save Yourself Some Grief, Try			
Pre-Research Counseling	2/3	April/81	2
Planning for Academic Computing	2/2	March/81	1
ICPSR - A University Wide Resource	2/2	March/81	7
2. ACCESS TO COMPUTING FACILITIES			
Accessing The Interactive			
Computing Systems	2/6	October/81	11
Getting Started on the HP	2/4	June/81	11
3. HARDWARE			
Microcomputer Policy	3/1	January/82	4
Using Micros As Terminals	2/7	Nov/Dec/81	5
TN Print Train Available	2/7	Nov/Dec/81	12
NTSU Computing Center General			
Equipment List	2/7	Nov/Dec/81	21
New Tape Drives	2/6	October/81	2
General Terminal Features	2/3	April/81	6
Telex Customer Phone Number	2/3	April/81	8
Take-Home Terminals Now			
Available For Faculty Use	2/2	March/81	7
Terminal Maintenance Contract	2/1	January/81	7
The New 3330/3350 Compatible Disk			
Drives - How "The Change" Can			
Effect You	1/3	October/80	6
News From College of Business			
Computing Center	1/3	October/80	12
Computer Acquisition Status Report	1/2	June/80	1
4. INPUT			
Procedure for Processing Class L Jobs	3/1	January/82	6
Revised Forms Description Guide	2/7	Nov/Dec/81	11
Guidelines For Tape Processing	2/6	October/81	4
Reading Disk Data Sets	2/6	October/81	5
Job Card Preparation	2/3	April/81	2
Disk Data Set Naming Conventions	2/2	March/81	2
Memory No Longer A Factor In Job			
Class Selection	1/3	October/80	10

<u>SUBJECT</u>	<u>VOL/NO</u>	<u>MONTH/YEAR</u>	<u>PG</u>
5. OUTPUT			
Priority Setting for Batch Jobs	2/4	June/81	2
Capturing Output from OSJR	2/4	June/81	6
Job Processing Schedule	2/3	April/81	4
Routing Jobs from OSJR	2/2	March/81	3
High Speed Print/Punch in Music	2/2	March/81	3
Secure Job Pickup	2/2	March/81	6
6. SERVICES AVAILABLE			
Data Entry Services	1/3	October/80	10
7. SOFTWARE			
A) General			
Software Acquisitions Since January 1981	3/1	January/82	6
Sorting Data Sets Efficiently	2/7	Nov/Dec/81	6
Software - What We Have And What We Support	2/6	October/81	18
Stat Users Beware	1/2	June/80	10
B) MUSIC			
PASCAL Documentation	3/2	February/82	10
Obtaining Help While Logged On To MUSIC	3/1	January/82	8
MUSIC Sort Facilities	3/1	January/82	9
CAT: An Additional Way to Browse Your Save Library Files	2/6	October/81	7
MUSIC Notes	2/6	October/81	11
Computer Assisted Instruction on SAS	2/6	October/81	12
MUSIC Version 5	2/5	September/81	5
Helpful Hints in Using the MIME-2A Tester	2/5	September/81	6
Electronic Mail System Revisited	2/5	September/81	7
Additional CAI Available Through MUSIC	2/4	June/81	4
Learning To Use MUSIC	2/4	June/81	6
MUSIC Notes	2/3	April/81	5
MUSIC Notes	2/3	April/81	7
MUSIC Notes	2/2	March/81	4
OSJE/OSJR	2/1	January/81	5
Management Of MUSIC SAVE Library Files	2/1	January/81	6
C) SPSS			
Using the MANOVA Procedure for Various Anova Designs	3/2	February/82	10
Printing "Very,Very Small" Numbers	3/1	January/82/13	
Undocumented MANOVA Conventions	3/1	January/82	13
Manipulating "Space" in SPSS	2/7	Nov/Dec/81	13
SPSS Enlightenments	2/7	Nov/Dec/81	14
Known Errors In SPSS Release 9	2/6	October/81	12
SPSS Statistical Algorithms Available	2/6	October/81	14
SPSS 9	2/5	September/81	9
D) SAS			
Problems: PROC TRANSPOSE, PROC CONVERT	3/2	February/82	14
SAS BUG	2/5	September/81	9
Using Multiple INFILE Statements	3/1	January/82	14
Using SAS/ETS PROC FORECAST	3/1	January/82	14
The Type I, II, III, and IV Dilemma	2/3	April/81	11
E) BMDP			
BMDP	3/1	January/82	14

<u>SUBJECT</u>	<u>VOL/NO</u>	<u>MONTH/YEAR</u>	<u>PG</u>
F) OTHER			
WATERLOO SCRIPT Available			
in Batch Mode	2/7	Nov/Dec/81	19
PRIME Statistical Package Available	2/7	Nov/Dec/81	20
Statistical Programs Acquired	2/7	Nov/Dec/81	20
GALILEO IV	2/2	March/81	7
ANOVA Models Tested by Various			
SPSS and SAS Options	2/1	January/81	10
Program Products Available For MVT	2/1	January/81	3
VM/370	1/2	June/80	7
Behavioral and Social Science Data			
Analysis Using Osiris	1/2	June/80	13
8. USAGE STATISTICS			
MUSIC Usage Statistics 1980-1981	3/1	January/81	10
AS/5000 Usage Statistics-September 1,			
1980 to July 31, 1981	2/5	September/81	11
9. UTILITY APPLICATIONS			
SETting The Attributes Of Your			
Save Library	3/1	January/82	7
#EXCHANGE: A New MUSIC Utility To			
Archive Save Library Files	2/6	October/81	8
New "Utility" Procedure	2/4	June/81	3
Changing Dataset Names With IEHPRGM	2/3	April/81	5
Deleting Disk Files	2/2	March/81	3

WORKSHOP/SEMINAR CHECKLIST

The following checklist was compiled to help the Academic Computing staff get a feeling for the needs of the users of academic computing facilities in terms of workshop/learning experiences. Please check all areas in which you think workshops are needed, and feel free to add as many other items as you deem necessary.

1. MUSIC

A. General overview (Introduction) ____

B. Working with Files ____

C. Accessing the OS operating system ____

D. Using the editors ____

E. MUSIC/Script (Introduction) ____

F. IIS ____

G. OTHER _____

2. HP/2000

A. General overview ____

B. Working with files ____

C. OTHER _____

3. Statistical Packages

A. SPSS

1. General Overview (Introduction) ____

2. OTHER _____

B. SAS

1. General Overview (Introduction) ____

2. OTHER _____

C. BMDP

1. General Overview (Introduction) ____

2. OTHER _____

D. OTHER Statistical Packages

1. Package _____

Presentation _____

2. Package _____
Presentation _____

3. Package _____
Presentation _____

4. JOB Control Language (JCL)
A. General overview _____
B. Specific Topic(s) _____

5. IBM Utilities
A. General Overview _____
B. Specific utilities/topic(s) _____

6. MISCELLANEOUS Topics (you list)

After completing the checklist, please return to:
Academic Computing Services
The Computing Center
NT Box 13495
North Texas State University
Denton, TX 76203

GET A "SUBSCRIPTION" TO BENCHMARKS

BENCHMARKS is a vital link between the NTSU Computing Center and the users of our facilities. It is important for all users of the computing facilities to maintain a file of these newsletters because they contain materials which will periodically update existing documents as well as information and suggestions on uses of OS-MVT, MUSIC, the HP-2000, and other resources available to NTSU students and faculty. To facilitate the dispersal of BENCHMARKS, *** FREE *** subscriptions are now available. To receive yours, send the following information to us either by "snail mail" (the post office or campus mail) or electronically, through the MAIL facility on MUSIC.

Name _____

Mailing Address _____

PLEASE GIVE A CAMPUS ADDRESS (NOT BOX) IF POSSIBLE - It's Cheaper !

PLEASE RETURN TO:

Academic Computing Services
The Computing Center
NT Box 13495
North Texas State University
Denton, TX 76203